Release of the 1997 Mature and Young Women Data

Researchers can now obtain data from the 1997 survey of mature and young women. These data include information gathered during the 18 interviews of the mature women and 19 interviews of the young women conducted since 1967 and 1968, respectively. The data are available on compact disc.

The 1997 survey collected information from 2,608 members of the mature women cohort, or 61.0 percent of the original respondents who are not known to be deceased (some respondents who have left the survey for other reasons may now be deceased). In the young women cohort, 3,049 respondents, or 61.1 percent of original respondents not known to be deceased, participated in the survey. If the retention rate is defined as the percentage of original respondents, including deceased respondents, who were interviewed in 1997, then the mature women's retention rate is 51.3 percent and the young women's retention rate is 59.1 percent.

As in 1995, the 1997 survey of the mature and young women was administered using a computer-assisted personal interview (CAPI). The two cohort groups were again surveyed during the same period with the same instrument, increasing the efficiency of the survey and making it easier for researchers to compare the cohorts. The CD-ROM presents the data for the two cohorts separately.

The 1997 survey generally collected the same core information as in previous years. Respondents provided extensive labor force and work history information and completed

modules on household member characteristics, education, health, income, assets, training, pensions, husband's work experiences, and geographic mobility.

Three main sets of questions from the 1995 survey were not included in 1997. In 1995, respondents in both cohorts answered questions about their responsibility for child care, updated basic demographic information on all of their biological and adopted children, and reported types of job discrimination experienced. These questions were all dropped in 1997 to make room for the transfers section.

One major question module, a section on parents and transfers similar to that included in the 1993 young women's survey, was added in 1997. This section begins with the collection of biographical and health information about the respondent's and her husband's parents. Information is then gathered about the parents' residences, income, assets, and debts. If the parents are deceased, respondents report whether there was a will and how the estate was divided. Finally, questions examine the nature and extent of transfers of time and money from respondents to their parents. Stepparents are included in the collection if they are currently married to a biological parent.

The 1997 mature women's survey also included a special collection of transfers information from respondents who had one or more daughters in the cohort of young women. This section asked the respondents to report the amount of help in the form of time and money which they had received from their daughters. Although the sample is not representative of all mothers and daughters, researchers can use these data to compare generational perceptions about the amounts of time and money transferred from the daughters to the mothers.

New editions of the *NLS of Mature Women User's Guide* and the *NLS of Young Women User's Guide* will soon be available. These cohort-specific guides update information found in

the previous guides through 1997 and include several new sections. They also explain how researchers can use the data more efficiently.

Data for the NLS of Mature Women (1967–97) and NLS of Young Women (1968–97) are distributed on a compact disc that contains the longitudinal record of each respondent. These data include answers to interview questions, edited and created variables based on these raw responses, basic geographic information provided by the Bureau of the Census, and data from a 1968 survey of schools attended by the young women. The disc also contains documentation files and search and retrieval software that enables users to easily peruse, select, and extract variables.

In addition, the CD-ROM includes data from the special 1989 mature women's pension plan data collection, as well as software designed for use with these variables. Using information gathered from the respondents on the names and addresses of the companies providing their pensions, Census Bureau staff members collected pension plan descriptions, which contain the formulas for calculating benefits for workers at different ages and with various years of service with the firm. Staff members at the Survey Research Center, University of Michigan, then coded this information into a standardized format to drive the "pension calculator," a program that computes how much money mature women respondents will receive in benefits upon retiring.

Supplemental documents, including a questionnaire, are distributed with each data set. Researchers interested in purchasing these data should contact NLS User Services (see back page for contact information).

NLSY97 Geographic Variables

To provide basic geographic information to all users, the main NLSY97 CD-ROM includes variables intended to permit researchers to identify key characteristics of the respondent's area without needing access to the geocode CD-ROM. These main NLSY97 geographic variables are described in the first section of this article. The next section then describes geographic variables available on the NLSY97 geocode CD-ROM.

Geographic variables on the main NLSY97 CD-ROM

Three variables describe the area where the youth lived at the survey date; the same characteristics are also provided for the area where the youth lived at age 12. The first variable classifies respondents as residing in one of four regions—Northeast, North Central, South, and West—defined by the Census Bureau (see the *NLSY97 User's Guide* for the specific States in each division).

Another variable reported in the NLSY97 main file describes whether respondents' survey date and age 12 residences are urban or rural. As defined using Census Bureau criteria, urban places are "closely settled, named communities that generally contain a mixture of residential, commercial, and retail areas, and have a population greater than 2,500."

The final geographic characteristic included in the main file concerns respondents' Metropolitan Statistical Area (MSA) and central city status at their survey date and age 12 residence. Using this information, researchers can identify whether the respondent lives in an MSA and whether that residence is in a central city.

Geographic variables on the NLSY97 geocode CD-ROM

The majority of geographic data are found on the NLSY97 geocode compact disc. These variables provide more detailed statistics for each respondent's county of residence. To acquire a geocode CD-ROM, researchers must go through a detailed application process and sign a confidentiality agreement before BLS will release the data. The remainder of this article describes the geocode variables and explains the confidentiality and accessing requirements for the geocode CD.

In addition to all main file data, the geocode CD provides a variety of statistics for the counties where respondents lived at the round 1 interview date. The first group of geocode variables lists the basic demographic information for respondents' counties. These data include land area in square miles; population by race, age, and gender; and birth and death rates. Another variable reports the percent of persons who lived in a different house or State in 1990 compared to their residence in 1985, providing information about migration rates for the respondent's area.

Factors that might influence the respondent's education and employment outcomes are the focus of several other geocode variables. These provide the number of serious crimes, households with children, female householders with no spouse present, persons with high school or college degrees, and families below the poverty level. A pair of variables summarizes medical availability for each county, reporting the number of active nonfederal physicians and community hospital beds.

Geocode variables associated with economic and labor force issues include the size of the county's civilian labor force, the percent employed in various industries, and the percent of workers age 16 and over with jobs outside their county of residence. Income variables include per capita money income for the respondent's county, per capita personal income, and median

family money income. The unemployment rates for metropolitan areas and for the portions of the State not in an MSA are also reported.

The final geocode variables involve colleges attended by respondents. Survey staff use information from the Integrated Postsecondary Education Data System (IPEDS) to provide users with the code identifying each college attended by the respondent and its location. These codes can be used to associate the NLSY97 respondent's college with various characteristics of the institution contained in the IPEDS database.

Table 1 summarizes the geographic variables available on the geocode CD-ROM and provides information about the source of data for each variable. Most variables are based on the 1994 *County and City Data Book (CCDB)* prepared by the Census Bureau. The *CCDB* data file includes information from the 1990 Census of the Population and Housing, the Current Population Surveys, and other supplemental data derived from a variety of federal government and private agencies.

Table 1. **Round 1 geocode variables and data sources**

Variable	Data source
GEO1, GEO2	County and state of residence (interview data)
GEO3–GEO5	MSA/CMSA/NECMA of residence and PMSA of residence (interview data and 1994 *CCDB)*
GEO6	Quality of match for geographic residence variables
GEO7	1990 land area in square miles (1994 *CCDB)*
GEO8	1992 population (1994 *CCDB)*
GEO9	1992 population, 1990 square miles (1994 *CCDB)*
GEO10–GEO15	1990 population by racial group (1994 *CCDB)*
GEO16–GEO25	1990 population by age group (1994 *CCDB)*
GEO26	1990 population—base for items GEO15 to GEO25 (1994 *CCDB)*
GEO27	1990 male population (1994 *CCDB)*
GEO28	1990 percent of persons 5 years and older living in different house in 1985 (1994 *CCDB)*
GEO29	1990 percent of persons 5 years and older living in different house, same state in 1985 (1994 *CCDB)*

GEO30	1990 percent of persons 5 years and older living in different house, different state in 1985 (1994 *CCDB)*
GEO31	1990 family households, percent with own children under 18 years (1994 *CCDB)*
GEO32	1990 female householders (no spouse present), family households (1994 *CCDB)*
GEO33	1990 female householders (no spouse present), family households, percent with own child (1994 *CCDB)*
GEO34–GEO36	1988 birth rate information (1994 *CCDB)*
GEO37	1988 population—base for GEO36 and GEO39
GEO38, GEO39	1988 death rate information (1994 *CCDB)*
GEO40	1988 deaths of infants under 1 year per 1,000 live birth (1994 *CCDB)*
GEO41	1988 active nonfederal physicians per 100,000 population [copyright] (1994 *CCDB)*
GEO42	1991 community hospital beds per 100,000 population [copyright] (1994 *CCDB)*
GEO43	1991 serious crimes per 100,000 population (1994 *CCDB)*
GEO44	1990 persons 25 years and over, percent high school graduate or higher (1994 *CCDB)*
GEO45	1990 persons 25 years and over, percent with bachelor's degree or higher (1994 *CCDB)*
GEO46	1989 median family money income (1994 *CCDB)*
GEO47	1989 per capita money income (1994 *CCDB)*
GEO48	1989 percent of families with income below poverty level (1994 *CCDB)*
GEO49	1990 total families - base for GEO48 (1994 *CCDB)*
GEO50	1989 percent of families with female householder (no spouse present) below poverty level (1994 *CCDB)*
GEO51	1990 female householders (no spouse present), family households—base for item in GEO50 (1994 *CCDB)*
GEO52	1989 percent of persons with income below poverty level (1994 *CCDB)*
GEO53	1989 percent of related children under 18 years below poverty level (1994 *CCDB)*
GEO54	1990 Workers 16 years and over, percent working outside county of residence (1994 *CCDB)*
GEO55	1991 civilian labor force (1994 *CCDB)*
GEO56	1991 civilian labor force—number unemployed (1994 *CCDB)*
GEO57	1991 civilian labor force—unemployment rate (1994 *CCDB)*
GEO58	1990 civilian labor force (1994 *CCDB)*
GEO59	1990 civilian labor force—percent female (1994 *CCDB)*
GEO60	1990 civilian labor force—percent unemployed (1994 *CCDB)*
GEO61	1990 civilian labor force—number employed (1994 *CCDB)*
GEO62–GEO67	1990 civilian labor force, percent employed various industry groups (1994 *CCDB)*
GEO68	1990 per capita personal income (1994 *CCDB)*
GEO69	1997 FICE code of 1st most recent college attended [IPEDS]
GEO70	1997 location of 1st most recent college attended—state [IPEDS]
GEO71	1998 continuous employment rate for labor market of current residence (May 1998 *Employment and Earnings)*

Confidentiality and accessing requirements

NLSY97 geographic data are provided voluntarily by respondents who have been assured that the data will only be used for research purposes and that results from that research will only be made public in summary or statistical form so that individuals cannot be identified. Because of the level of detail of these geocode data, accessing requirements are used to further protect respondent identities.

If a researcher should violate the agreement with BLS, the consequences to the entire research community would be much more severe than the individual punishment received. Any released information that could potentially be used to identify individuals might damage the trust that respondents place in the pledge of confidentiality from BLS. If respondents should feel that their privacy is threatened, participation in the surveys will drop and the quality of data will decline. Furthermore, any violation of the accessing agreement could jeopardize future releases of geographic information. The continued availability and quality of this important data set depends upon compliance with these BLS standards.

Researchers who wish to use the geocode variables must first agree to not disclose any sensitive respondent information and must explain the security procedures their institution will have in place to protect the data. In their published work, researchers must not include any information that could lead to identification of a respondent. At the end of the research project, the user is required to return the geocode disc to BLS; this further ensures respondent confidentiality.

The accessing agreement process begins when researchers file an application with BLS. This application requests information about the topic of research and need for geographic data, asks researchers to describe security procedures, and requires researchers to sign a non-

disclosure agreement. To receive more information or an application, contact NLS User Services or Rita Jain at BLS (see the back cover for contact numbers).

Behind the Scenes: NLSY97 Interviewer Training

Efficient, professional interviewing is a key element of any survey effort. It is essential that the interviewers conduct the interviews in a consistent manner with each respondent to avoid introducing bias into the data. Further, because the NLSY97 interviews are conducted using laptop computers, interviewers must be very comfortable with the computerized survey program.

With these considerations in mind, NLSY97 survey staff implement a training program designed to equip interviewers with the interviewing and computer skills necessary to collect high-quality data. Because many researchers may be unaware of this important aspect of the NLSY97 survey, this article describes the basic components of the interviewer training program used in round 2: The training of field managers, at-home interviewer training, and in-person interviewer training.

Field manager training

The first step in the NLSY97 training program is the instruction of field managers in both interviewing and management techniques. Selected from the group of interviewers who have previously worked on either the NLSY97 or NLSY79, field managers are responsible for overseeing all interviewers within their geographic area. Field managers monitor production and quality control, plan field strategies, and provide a communications link between interviewers

and the central office. They also assist interviewers in gaining the cooperation of reluctant respondents and in locating hard-to-find respondents.

Since the field managers are experienced NLSY97 or NLSY79 interviewers, they first complete the at-home interviewer training described in the next section. This enables them to become familiar with the content of the survey and the operation of the computerized interview program. Field managers then attend a 2½-day management seminar at which they expand their management skills, share ideas, develop problem solving strategies, and generate an optimal plan to achieve success during the field period.

At-home interviewer training

The interviewing staff for the NLSY97 normally includes a large number of interviewers with experience administering either the NLSY97 or NLSY79. In round 2, for example, approximately 75 percent of the interviewers were experienced with the NLS. To minimize costs while maintaining a high-quality corps of interviewers, NLSY97 survey staff have developed an at-home training course for these experienced interviewers.

The at-home training course guides interviewers through the entire survey process, including getting interviewing assignments, locating and contacting the respondent, gaining cooperation, conducting the interview, and following the post-interview administrative and reporting procedures. This course consists of a 16-hour self-study program of readings, a video, practice interviews, and written exercises. The at-home training program also features a computerized tutorial that guides interviewers through the various functions of the laptop survey program. The field managers, as well as technical support services staff, are available to address questions and problems throughout the training period.

After completing the at-home course, trainees are required to conduct a checkout practice interview with their field manager to demonstrate mastery of the computerized questionnaire. This practice interview session also allows the trainees to ask any remaining questions about the other parts of the interview process.

In-person interviewer training

Interviewer training for new interviewers consists of a 4-hour self-study and a 3½-day in-person training program. Much like the at-home training for experienced interviewers, the self-study activities include a computerized tutorial, readings, written exercises, and a video. This part of the training is designed to provide the new field interviewers with background knowledge about the survey and the computer before they attend the in-person training.

When they arrive at the in-person training site, interviewers are divided into small groups of approximately 12 trainees each. Each of these groups is led by two trainers. NLSY97 survey staff monitor all the training groups to ensure consistency among interviewers.

A major component of the in-person training focuses on NLSY97 interviewing procedures. Conducting the survey involves a number of steps that must be completed by interviewers both before and after the actual interview takes place, and consistent completion of these administrative tasks is important to maintaining high-quality data. In sessions throughout the training period, interviewers review the purpose of the survey, the content of the questionnaire, and the operation of the computer. Trainees also receive in-depth instruction in locating respondents, gaining permission to conduct an interview, and persuading reluctant respondents to complete the survey. In the final sessions on survey procedures, interviewers learn how to perform important administrative tasks such as transmitting the data to the central office and reporting to the field manager about the status of the case.

A main part of the in-person training focuses on the actual administration of the survey. During each day of the training period, interviewers conduct practice interviews to become familiar with the content of the survey and the operation of the computer. In some practice interviews, trainees administer the entire survey. In other sessions, trainees perform only partial interviews so that they can focus on the three most complicated sections: The PIAT Math Test, the employment section, and the schooling section. These practice interviews are intended to give all trainees as much opportunity as possible to administer the survey, a frequent request from interviewers after previous CAPI training.

The last half-day of training is reserved for a scenario exercise. This exercise requires trainees to walk through each step of the interviewing process from beginning to end and allows them to conduct a complete practice interview. In addition to this final practice scenario, new interviewers also complete a check-out practice interview with their field manager by telephone.

Training does not end when the field period begins. After all interviewers have a chance to contact and interview at least one respondent, field managers hold group conference calls so that the interviewers can share their experiences and concerns. These calls also help interviewers hone techniques for locating hard-to-find respondents and for convincing respondents to participate in the survey. It is particularly important for interviewers to continually improve these refusal conversion and locating skills, because they are a key part of obtaining a high overall response rate.

Throughout the field period, field managers continue to communicate with their interviewers about common problems or any other issues that arise. Interviewers communicate through field memos, e-mail, and weekly reporting calls with the field manager. This constant communication ensures that all interviewers will be made aware of any changes in field

techniques or corrections to the instrument that may occur during the field period. It also allows field managers to identify interviewers who are having difficulty and assist them in improving their techniques before any real problems develop.

The final part of the training process concerns the trainers rather than the trainees. A successful field effort requires constant review and improvement of the training process. At the end of the training, debriefing meetings are attended by all levels of project staff. Topics discussed include all areas of at-home and in-person training, and staff members are invited to contribute any suggestions that may help prepare field staff for successful data collection or improve future training programs. For example, past debriefing meetings have resulted in additional computer-based tutorials, which have proven beneficial to the interviewers, and additional practice interviews. These steps allow NLSY97 staff to ensure continued high response rates and quality data for the research community.

NLSY79 Income and Wealth Data

Although many surveys collect information about a respondent's income, the NLSY79 goes further by obtaining information on the income, assets, and debts of the respondent. As a result, wealth, which is equal to a respondent's assets minus his or her debts, can be calculated providing another dimension of the respondent's financial resources. Researchers can also more precisely measure wealth for low-income and minority households because the original NLSY79 panel contained a supplemental sample of 5,295 blacks, Hispanics, and economically disadvantaged non-black/non-Hispanics (the economically disadvantaged non-black/non-

Hispanics were dropped from interviewing after the 1990 survey). This article examines the sets of questions relating to income and wealth and summarizes data quality research on the wealth information.

Although this article focuses on the NLSY79 cohort, researchers should be aware that respondents in each of the other NLS surveys have also answered questions about their income and wealth on a regular basis. These data permit comparison of incomes, assets, and debts in several generations of U.S. residents. Interested researchers should consult the *NLS Handbook* or the cohort-specific *NLS User's Guides* for details about the types of information available.

Income

The NLSY79 income section has remained fairly constant across all of the survey years, allowing researchers to make consistent comparisons over time. This section collects detailed information about the income of respondents and their spouses. In addition, surveys in earlier rounds gathered some basic information about the income of opposite sex partners; this collection has grown more detailed in recent interviews.

In 1979-82, different questions were asked based on whether the respondent was classified as independent. Independent respondents were those who were 18 years of age or older, had a child, were enrolled in college, were married, or were living outside their parents' home. These independent respondents answered the same questions in 1979-82 as described in the next paragraph for the 1983 survey. Those respondents who did not meet any of these criteria were asked only to report their income from wages, salaries, and tips; unemployment compensation; housing subsidies; and all other sources combined.

Every survey beginning with 1983 has asked all respondents about their and their spouses' income from each of the following sources for the previous calendar year:

- Military income
- Wages, salary, tips
- Net business income
- Net farm income
- Unemployment compensation

Beginning in 1990 respondents reported the income of their opposite sex partners from each of the first four categories above. Unemployment compensation was added to this series in 1994.

The survey then collects information about income received jointly by the respondent and his or her spouse from a number of sources listed below. If the information was not collected in all surveys, the years in which the source has been included are listed in parentheses. As with the income sources described above, these data were only collected from respondents who met certain independence criteria in the 1979-82 surveys:

- Child support (1982-98)
- Alimony (combined with child support 1982-93; in "other" category 1994-98)
- AFDC payments
- Food stamps
- Other welfare and SSI
- Education benefits or grants
- Disability benefits (1980-98)
- Veteran's benefits (1980-98)
- Other (interest, dividends, rent)
- Housing subsidy

Beginning in 1994, respondents were asked about income from the above sources (except subsidized housing), that was jointly received by the respondent and his or her opposite sex partner.

A few additional income questions have been asked in some survey years. The 1983-88, 1996, and 1998 surveys asked about financial support respondents and their spouses received from their parents. Beginning in 1988, the interview included a question about whether the respondent had received any money from inheritances or gifts in the past calendar year.

Before the collection of detailed opposite sex partner information began, the survey asked respondents to provide some basic information about partner income. The respondent first answered a series of “yes” or “no” questions about sources of the partner’s income and then reported the partner’s total income from all sources combined.

Finally, in all surveys, the total income received by all other household members combined is recorded, along with a list indicating the sources of this income. In 1979-86, the respondent’s parent usually provided this information if the respondent was still residing in the parental home. In all other cases, it was provided by the respondent.

The reference period for most income questions is the past calendar year. Users should note that the switch to a biennial survey schedule means that, beginning in 1994, income data are no longer available for every year even though the questions are still included in every survey. That is, the 1994 survey referred to the respondent’s 1993 income, the 1996 survey collected information about 1995 income, and the 1998 interview asked about 1997 income. As a result, no data are available for the 1994, 1996, or 1998 calendar years.

Wealth

Wealth, or a respondent’s assets minus his or her debts, provides a much more complete picture of the respondent’s financial resources than information on income alone. The longitudinal nature of the NLSY79 makes it a valuable source for this type of information, because few surveys measure a respondent’s assets and debts over time.

The early survey years included a few questions on assets; no information was collected on debts. In 1979-84, asset questions were addressed only to respondents who were classified as independent. While these questions are somewhat limited, they do permit researchers to

determine when saving begins, how savings habits are formed, and how persistent savings habits are.

Specific questions in 1979-82 collected data about whether the respondents (and their spouses, if applicable) owned or were purchasing their home or apartment, whether they owned any vehicles, and whether they had any savings. However, the survey did not ask respondents to provide information about the value of their home, vehicles, or savings account or the amount of their mortgage. In 1983 and 1984, the only asset question included in the survey asked respondents whether they owned their home or apartment.

Beginning in 1985, when all respondents had reached age 18, the NLSY79 included a much larger wealth section. Core questions on assets and debts have remained fairly constant since that time, although the wealth section has grown in length and detail as respondents have aged. The major exception is the 1991 survey, which did not include any asset and debt questions.

Respondents report their wealth in two-stage questions. First, the survey asks whether the respondent and his or her spouse own a particular asset or have a particular debt; if so, the respondent is asked to report the value of the asset or the amount of the debt. The 1985-90 and 1992-98 surveys collected information about the following:

- Ownership of home or apartment; market value and amount owed on property in mortgages and other debts
- Ownership of farm property, a business or professional practice, or other real estate; market value and amount owed
- Ownership of vehicles for personal use; market value and amount owed
- Ownership of any other items worth over $500 each, such as furniture, electronics, jewelry, or collectible items; market value
- Whether the respondent and spouse have any other debts of over $500 each; amount owed

Surveys from 1985 to 1987 also asked a single question about whether the respondent and his or her spouse had any money assets. These assets included

- Savings or checking accounts
- Money in a savings and loan, money market fund, or credit union
- U.S. savings bonds
- An IRA or Keogh account
- Certificates of deposit
- Stocks, bonds, or mutual fund accounts
- Rights to an estate or trust
- Loans or mortgages owed to the respondent.

In 1988, two new question sets were asked, recording the ownership and value of stocks, bonds and mutual fund accounts and the ownership and value of rights to an estate or trust. In 1994, the money assets question was further divided, and respondents were asked separately about (1) IRA or Keogh accounts, (2) tax-deferred annuities (such as a 401k or 403b), and (3) any certificates of deposit or money owed to the respondent for personal loans or mortgages.

Beginning in 1990, the survey added a set of questions which gauge the respondent's financial situation as a whole. Respondents are asked to calculate whether they would come out ahead, break even, or still be in debt if they sold all their assets and paid off all of their debts. If the respondent would have money left over or be in debt, he or she is asked to report the amount.

Finally, there are a few assets questions that have been asked only once. In 1985, respondents reported the make, model, and year of their vehicles if they did not know the value. The 1988 survey asked respondents whether the amount of their savings had changed since the last interview and, if so, by what amount.

Topcoding. Because the NLSY79 is a public use data set distributed widely throughout the research and public policy communities, the survey takes extensive measures to protect the confidentiality of respondents. One method of ensuring confidentiality is to "top code" unusually high values.

The NLSY79 has used several different top coding methods for income and wealth data. From 1979 to 1984, every NLSY79 income question that elicited a response above $75,000 was truncated to $75,001. From 1985 to 1988, the cutoff values for truncation were increased to $100,000 and $100,001, respectively. Assets were truncated in much the same way during the 1979-88 period; the cutoff value varies across asset categories. Unfortunately, this algorithm results in a sharp downward bias in the mean value of NLSY79 income holdings since the entire right hand tail is truncated, although the median remains unaffected.

To address this problem, a new method was implemented for both income and assets beginning in 1989. The new top code approach replaces all values above a designated cutoff with the average of all outlying values. Using this method ensures that more accurate means and other statistics can be calculated.

The top coding method was changed slightly in 1996. Rather than using a designated cutoff point, survey staff identified the top 2 percent of respondents with valid values. All the amounts within that top range were averaged and that averaged value replaced all reported amounts in the top 2 percent. For some variables, this change reduces the number of respondents with top coded information, permitting more detailed and accurate analysis.

Data quality

Two researchers have recently examined the quality of the NLSY79 wealth data. As part of the effort to redesign the NLSY79 survey for future rounds, Engelhardt (1998) compares NLSY79

wealth data to information collected in the Surveys of Income and Program Participation (SIPP). He finds that the mean and median asset values are fairly similar across asset categories for all age, marital, and racial groups, although rates of asset ownership are slightly lower in the NLSY79. Engelhardt also concludes that item nonresponse for wealth questions is lower among NLSY79 respondents than in several other large national surveys with significant wealth components, suggesting that the NLSY79 may be less prone to errors in measuring wealth due to nonresponse. However, Engelhardt cautions that this may be due to the relatively young ages and simple portfolios of the NLSY79 respondents and that nonresponse may rise significantly in the future.

Zagorsky (1997, 1999) also examines the quality of the NLSY79 wealth data. Like Engelhardt, he finds that nonresponse is fairly low and that NLSY79 wealth data are roughly similar to other national financial surveys, such as the Survey of Consumer Finances (SCF), the Panel Study of Income Dynamics (PSID), and the Survey of Income and Program Participation (SIPP). However, Zagorsky concludes that NLSY79 respondents hold slightly more valuable assets than comparable respondents in other surveys. This slight upward bias may be due either to nonrandom attrition or to a greater likelihood to respond honestly to a familiar interviewer over the course of a long-running survey. Finally, Zagorsky finds that respondents tend to underestimate their net worth when answering the 1990–98 questions about selling all assets and paying all debts.

Although a number of research articles have been written using the NLSY79 income data, relatively little research is based on the wealth data. Users can identify current research in these areas by searching the NLS online bibliography at **http://www.chrr.ohio-state.edu/nls-bib/** for words such as “income,” “assets,” “debts,” and “wealth.”

For more information about wealth data in the NLSY79, interested researchers should consult the *NLSY79 User's Guide* or the *NLS Handbook*. Users can locate the questions described in this article by looking in the income and assets sections of the various questionnaires. The NLSY79 CD-ROM contains the data in the INCOME, ASSETS, and MXXVAR record types. The documentation items and the NLSY79 data CD-ROM can be obtained from NLS User Services (contact information is provided on the back on this newsletter).

References

Engelhardt, Gary V. "Income and Wealth in the NLSY79." Presented at the NLSY79 Redesign Conference, September 1998, Washington, D.C. gary.v.engelhardt@dartmouth.edu

Zagorsky, Jay L. "The NLSY79 Wealth Data Evaluation." Center for Human Resource Research, The Ohio State University. Columbus, OH: 1997.

Zagorsky, Jay L. "Young Baby Boomers' Wealth." *Review of Income and Wealth* Series 45, No. 2, pp. 135-156, June 1999.

Corrections to NLSY97 Data

The following data and documentation files are all available on the NLS User Services ftp site. Researchers should contact NLS User Services (see the back of this newsletter) for more information on obtaining these files.

A review of the NLSY97 data following the round 1 data release revealed a number of problems with the relationship codes in the household roster (HHI2). Survey staff have updated these relationship codes, as well as the associated gender variables on the HHI2 roster, and they are now available to the public. These updated data are contained in an ASCII file that includes the following variables: PUBID, 130 of the relationship variables found on the HHI2 roster, all HHI2 roster gender variables, the relationship of the responding parent to the youth respondent, and the responding parent ID variable. A documentation file provides information about the layout and use of the ASCII data file.

The review of the data also revealed that in 38 cases, either the respondents were not age-eligible for the NLSY97 cohort or the case was a duplicate record. An ASCII data file on the ftp site lists the identification numbers for the 38 dropped cases. These cases will be removed from the round 2 data release.

Finally, also available from NLS User Services is information about the occupations of NLSY97 youths reporting freelance employment. During the interview, youths described the type of work they performed; these data were then coded by survey staff into categories such as babysitting, mowing, pet care, and snow shoveling. Due to the length of time required for the coding process, these data were not available at the time of the round 1 release but can now be merged into the data file by interested researchers.

Frequently Asked Questions

NLS User Services encourages researchers to contact them with questions and problems they have encountered while accessing and using NLS data and or documentation. Every effort is made to answer these inquiries. Some recently asked questions that may be of general interest to NLS users are listed below with their answers. All of these questions refer to the recently released NLSY97 data.

Q1: I'm trying to find a race variable for the youth in the NLSY97. Is there a variable that reports whether a respondent is black, Hispanic, or non-black/non-Hispanic? I found the ethnicity variable, which identifies respondents as Hispanic, but these respondents fall into all categories in the race variable.

A1: The first thing to remember is that a person can be of Hispanic ethnicity and still be of any race. The questionnaire first asks whether respondents consider themselves to be Hispanic; it then asks all respondents, regardless of ethnicity, to identify their racial group.

During the interview, questions about race and ethnicity are asked only once, during the extended screener portion of the screener, household roster, and non-resident roster questionnaire. The answers to these questions are contained in the household roster, or HHI2, variables (R10803.–R11636.). For people in the household who are NLSY97 respondents, race and ethnicity information is then transferred into the youth questionnaire in two variables called symbols. Race information is summarized in the symbol variable KEY!RACE (R05387.) and ethnicity is identified in KEY!ETHNICITY (R05386.).

Assuming that you want Hispanic ethnicity to take priority over any racial identity, you can create a single race/ethnicity variable by first coding all Hispanic respondents using KEY!ETHNICITY. Then use KEY!RACE to code all the remaining non-Hispanic respondents as black, white, American Indian/Eskimo/Aleut, Asian/Pacific Islander, or other.

Users who examine the KEY!RACE variable will note that more than 1,000 respondents are coded as "something else." Researchers may be able to obtain more information about the race of these respondents by examining the respondent's race variable in the HHI2 roster. (The position of the respondent on the household roster can be identified by using the variable YOUTH_HHID.01, which is R05332.) In these variables, taken from the household roster, some respondents answered "other" and then specified that they were either Hispanic or of mixed race. When KEY!RACE was created, these answers were included in the "other" category. After the interview, survey staff hand-coded these "other specify" responses into Hispanic, mixed race, and other categories. The HHI2 data were adjusted accordingly but the KEY!RACE variable was left as it was during the interview.

Q2: We're trying to locate the variables indicating the youth's residential parents' educational attainment, age, marital status, and relationship to the youth. The survey asks about education in some screener items, but these variables are not on the CD. Further, we cannot find any variables indicating the marital status of the youth's biological parents and/or whether the youth's residential parents are both biological parents or one is a stepparent. We found a check in the parent questionnaire for the accuracy of the information from the screener, but not an actual variable telling us the parents' marital status.

Also, we have found screener variables indicating the educational status of non-household members, but we don't know who these people are. For example, are any of the people on the roster non-residential biological parents? How can we best obtain this information?

A2: Screener questions do not collect information for parents separately from other household members and non-resident relatives. Therefore, researchers must use the set of ID variables (e.g., YOUTH_MOMID, YOUTH_ADOPDADID, YOUTH_STEPMOMID, etc.) to identify the parents' line numbers on the household roster. Similarly, use YOUTH_NRMOMID and YOUTH_NRDADID to identify the line numbers of non-resident biological parents on the non-resident roster. These variables can be easily found by searching for "ID" using the any word search function.

Note that the relationship data have been corrected since the initial round 1 data release (see announcement above); in some cases, this may mean that the person identified in the parent ID variables actually has some other relationship to the youth. Researchers can use the updated relationship variables to double check the accuracy of the parent ID variables. These parent ID variables may be updated at a future date to reflect corrected relationship data; contact NLS User Services for details.

Once you have identified the line numbers of the people you are interested in, you can look at the rosters called HHI2 (for household residents, R10803.-R11636.) or NONHHI (for non-residents, R11637.-R11929.). These rosters contain information from the screener questions about birth date, education, employment, and marital status; you will have to loop through and create variables for the youth's parents from these rosters. Some parent information may be corrected in the PARHHI roster (R06945.–R07313.) in the parent questionnaire; users may wish to check these variables as well to ensure that they have the most up-to-date data.

Researchers can obtain more information about the SAS programming code required for this process by contacting NLS User Services.

Q3: I generated two SPSS frequency distributions for a marijuana use variable, each using one of the weight variables on the NLSY97 CD-ROM (question names SAMPLING_WEIGHT and CS_SAMPLING_WEIGHT). The differences between the two estimates are not dramatic; the percentages are about the same but the counts have some discrepancies. Can you explain why this happens and what the difference is between the two weight variables?

A3: The NLSY97 sample is actually composed of two groups, a set of respondents who represent a cross-section of the United States and an oversample of respondents who are black or Hispanic. For more information on sample selection, refer to chapter 3 of the *NLSY97 User's Guide*.

If you use the regular sampling weights (SAMPLING_WEIGHT, R12361.), you are analyzing all NLSY97 respondents. The weights (when divided by 100) will add up to an estimate of the number of U.S. residents in the sample age range in 1997. When you use the cross-sectional weights (CS_SAMPLING_WEIGHT, R12362.), all oversample cases are dropped because each oversample case has a zero weight in this series. The weighted results are similar because these weights are also designed to produce an estimate of the number of U.S. residents in the sample age range. Since there are fewer respondents if the oversample is omitted, each black or Hispanic respondent in the cross-sectional sample is given a larger value.

For research that includes analysis by race, you should probably ignore the cross-sectional weights. If you use these weights, you are eliminating a large number of black and Hispanic youths. As a result, your black and Hispanic estimates will be similar but statistically

measured with much less precision than if you used the regular sampling weights. If you never want to break your results down by race or your research focuses only on whites, then using the cross-sectional weights will save processing time and speed up your data analysis. However, very few users want to limit their analyses in this way.

Q4: How can I identify the CPS employer for NLSY97 respondents? (Note: The CPS employer is a term used in the NLSY79 survey to identify the respondent's current or most recent employer using methods similar to the Current Population Survey. Researchers sometimes limit their analysis to only the CPS employer instead of considering all employers.)

A4: Unlike the NLSY79, the NLSY97 does not ask specific questions that refer to the CPS employer; all employers are discussed in the employment section of the questionnaire. However, there is information available that can be used to determine which job or jobs are current for a given respondent.

First, you need a little information about the way employment information is collected during the interview. The respondent is asked to provide the names of all the employers (including family businesses) for whom he or she has worked since age 14. Then, in the YEMP-1800.xx variables, each employer is assigned a number (e.g., 9701, 9702, and so on through 9707) in the order in which they were reported by the youth.

After the employers are assigned a number between 9701 and 9707, the respondent reports the dates he or she started and stopped working for each employer. (These questions are not represented on the CD exactly as asked; they are reordered and reported in YEMP_STARTDATE.xx and YEMP_STOPDATE.xx.) At this point, the survey program sorts the jobs so that the most recent employer is employer #01, the next most recent is employer #02,

and so on. Throughout the rest of the employment section, the employer numbers remain constant, so that each variable containing the phrase “Job #01” or “Employer #01” refers to the same employer for a given respondent. However, “Employer #01” will not necessarily be employer number 9701. The variables titled YEMP_UID.xx provide a crosswalk between the two systems of identification.

At this point, you can use the current status flag (YEMP_CURFLAG.xx) to determine whether the respondent is currently employed at each of the jobs listed in his or her employer roster. If the respondent is not currently employed at any job, you can find the most recent job by examining the stopdates. If the respondent is currently employed at more than one job, you can look at the usual hours worked variables to determine which job has more hours. (In the NLSY79, the job with more hours is classified as the CPS job if the respondent has more than one job at the same time.)

If you prefer, you can also use the created event history variables to examine the respondent’s employment. For each week from the date the respondent turned 14 through the interview date, these variables provide the ID number (9701–9707) for each employer the respondent worked for in that week. The first job is reported in the EMP_STATUS array; dual jobs are recorded in the EMP_DUAL arrays. Again, you can use the YEMP_UID.xx variables to link the employer(s) in a given week with the information the respondent provided during the interview.

Completed NLS Research

The following list of recent research, based on data from the various NLS cohorts, has not appeared in its current form in a previous issue of *NLS News*. See the NLS Annotated Bibliography, located online at **http://www.chrr.ohio-state.edu/nls-bib/** for a comprehensive research list.

Averett, Susan L. and Korenman, Sanders D. "Black-White Differences in Social and Economic Consequences of Obesity." *International Journal of Obesity* Vol. 23, No. 2, pp. 166-173, February 1999. [NLSY79]

Bartel, Ann P. and Sicherman, Nachum. "Technological Change and Wages: An Interindustry Analysis." *Journal of Political Economy* Vol. 107, No. 2, pp. 285-325, March 1999. [NLSY79]

Caputo, Richard K. "Grandmothers and Coresident Grandchildren." *Families in Society: The Journal of Contemporary Human Services* Vol. 80, No. 2, pp. 120-126, March-April 1999. [Mature Women]

Caputo, Richard K. "Head Start, Poor Children, and Their Families." *Journal of Poverty* Vol. 2, No. 2, pp. 1-22, Summer 1998. [NLSY79, NLSY79 Children]

Gamoran, Adam; Mare, Robert D.; and Bethke, Lynne. "Effects of Nonmaternal Child Care on Inequality in Cognitive Skills." Discussion Paper #1186-99, Institute for Research on Poverty, University of Wisconsin-Madison, March 1999. [NLSY79, NLSY79 Children]

Gorman, E.H. “Bringing Home the Bacon: Marital Allocation of Income-Earning Responsibility, Job Shifts, and Men’s Wages.” *Journal of Marriage and the Family* Vol. 61, No. 1, pp. 110-122, February 1999. [NLSY79]

Handcock, Mark S.; Morris, Martina; and Bernhardt, Annette. “Trends in Earnings Dispersion in the National Longitudinal Survey of Youth and the Current Population Survey.” Working Paper 98-14, Population Research Institute, The Pennsylvania State University, April 1999. [NLSY79]

Hausman, Patricia. “On the Rarity of Mathematically and Mechanically Gifted Females: A Life History Analysis.” Ph.D. Dissertation, The Fielding Institute, May 1999. [NLSY79]

Heymann, S. Jody and Earle, Alison. The Impact of Welfare Reform on Parents’ Ability to Care for their Children’s Health. *American Journal of Public Health*, Vol. 89, No. 4, pp. 502-505, April 1999. [NLSY79]

Hight, Joseph E. “Young Worker Participation in Post-School Education and Training.” *Monthly Labor Review*, Bureau of Labor Statistics, Vol. 121, No. 6, pp. 14-21, June 1998. [NLSY79]

Leigh, Duane E. and Gifford, Kirk D. “Workplace Transformation and Worker Upskilling: The Perspective of Individual Workers.” *Industrial Relations* Vol. 38, No. 2, pp. 174-191, April 1999. [NLSY79]

Light, Audrey. “Estimating Returns to Schooling: When Does the Career Begin?” *Economics of Education Review* Vol. 17, No. 1, pp. 31-45, February 1998. [NLSY79]

Ludwig, Jens. "Information and Inner City Educational Attainment." *Economics of Education Review* Vol. 18, No. 1, pp. 17-30, February 1999. [NLSY79]

Lynch, Lisa M. "A Needs Analysis of Training Data: What Do We Want, What Do We Have, Can We Ever Get It?" *Labor Statistics Measurement Issues*, NBER Studies in Income and Wealth, Vol. 60. Eds. John Haltiwanger, Marilyn Manser, and Robert Topel. Chicago: University of Chicago Press, 1998. 405-426. [NLSY79, Young Men, Young Women]

Nixon, Lucia A. and Robinson, Michael D. "The Educational Attainment of Young Women: Role Model Effects of Female High School Faculty." *Demography* Vol. 36, No. 2, pp. 185-194, May 1999. [NLSY79]

Pavalko, Eliza K. and Smith, Brad. "The Rhythm of Work: Health Effects of Women's Work Dynamics." *Social Forces* Vol. 77, No. 3, pp. 1141-1162, March 1999. [Mature Women]

Payne, Allison Ann. "Child-Rearing, Self-Control, and Deviance: An Examination of Gottfredson and Hirschi's General Theory of Crime." Master of Arts Thesis, University of Maryland, 1999. [NLSY79, NLSY79 Children]

Thorpe, K.E. and Florence, Curtis Samuel. "Health Insurance among Children: The Role of Expanded Medicaid Coverage." *Inquiry-The Journal of Health Care Organization Provision and Financing* Vol. 35, No. 4, pp. 369-379, Winter 1998. [NLSY79]

Valentine, Sean and Mosley, Gordon. "Aversion to Women who Work and Perceived Discrimination among Euro-Americans and Mexican-Americans." *Perceptual and Motor Skills* Vol. 86, pp. 1027-1033, 1998. [NLSY79]

Veum, Jonathan R. "Training, Wages, and the Human Capital Model." *Southern Economic Journal* Vol. 65, No. 3, pp. 526-538, January 1999. [NLSY79]

Western, Bruce and Beckett, Katherine. "How Unregulated is the U.S. Labor Market? The Penal System as a Labor Market Institution." *American Journal of Sociology* Vol. 104, No. 4, pp. 1030-1060, January 1999. [NLSY79]

Wunnava, Phanindra V. and Ewing, Bradley T. "Union-Nonunion Differentials and Establishment Size: Evidence from the NLSY." *Journal of Labor Research* Vol. 20, No. 2, pp. 177-183, Spring 1999. [NLSY79]

Yoshikawa, Hirokazu. "Welfare Dynamics, Support Services, Mothers' Earnings, and Child Cognitive Development: Implications for Contemporary Welfare Reform." *Child Development* Vol. 70, No. 3, pp. 779-801, May/June 1999. [NLSY79, NLSY79 Children]

Zagorsky, Jay L. "Young Baby Boomers' Wealth." *Review of Income and Wealth* Series 45, No. 2, pp. 135-156, June 1999. [NLSY79]

Are You Working With NLS Data?

If you are, we are interested in your work!

- Have you received funding to sponsor a project using NLS data?

- Are you working on a paper that uses NLS data?
- Have you published a recent paper using NLS data?

If you have received funding on a project, are working on a paper, or published a recent paper that uses NLS data, please contact: User Services, Center for Human Resource Research, 921 Chatham Lane, Suite 100, Columbus, OH 43221 (614) 442-7300, e-mail: **usersvc@pewter.chrr.ohio-state.edu**